The need for digital certificates in 5G networks

White paper

The open and dynamic nature of 5G networks expands the attack surface and makes it harder to secure data in motion. Encryption using digital certificates is vital to mounting an effective defense. This paper examines the reasons why digital certificates are essential to 5G security and sets out some general principles for deployment.

Contents

What's driving the need for digital certificates

The dynamic nature of 5G networks and the opening up of the signaling layer to enable new services has grown the attack surface and made it more imperative for communications service providers (CSPs) to secure data-in-motion or data-at-rest. That's made security mechanisms such as encryption more critical — including the use of digital certificates to authenticate the sources and recipients of data.

CSPs have used certificates in relatively limited ways up to 5G, mainly to authenticate base stations or encrypt traffic and IPSec as described by 3GPP (and in some cases of 4G, the use of IPSec has been mistaken as optional).

In 5G, the need for digital certificates extends not only to base stations and data traffic but also across the application level so that the various applications deployed can also communicate securely. And the complexity doesn't stop there, because those applications are built on other layers of the network, right down to the virtualization layer where communication among containers and Kubernetes also needs to be secured.

In other words, the need for certificate-based security is distributed throughout the network in 5G, all the way to the virtualized instantiation of the network foundation itself.

Understandably, this new, expanded role for digital certificates has many CSPs feeling a little outside of their comfort zone. This paper takes a look at the reasons why these certificates are such an essential part of the security mix in 5G networks and outlines some general principles for deployment.

All-new security for a whole new "G"

Cloud-based infrastructures and agile-concentric networks are integral to 5G success, allowing mobile CSPs to deploy new services rapidly. They make it possible for CSPs to become DSPs — digital service providers — using slicing technologies that let multiple customers enjoy dedicated services on shared infrastructure. Automation is essential in this kind of complex, highly orchestrated environment, which presents new security challenges because it's out of step with rigid, inflexible classical security concepts.

At the same time, data traffic volumes continue to climb. So do the number and variety of connected devices, including smartphones and tablets on the consumer side and Internet of Things (IoT) devices that rely on machine-to-machine communication in industrial and other contexts.

The flat, open nature of today's networks poses an even greater security challenge, with users' online behavior and reliance on mobile apps exposing networks to DDoS, DoS and GTP protocol anomaly attacks.

All these conditions combined put at risk the reputation for strong security that CSPs established with 2G and 3G networks.

The role of digital certificates in the 5G environment

Encryption and authentication were introduced through 3GPP standardization in response to lessons learned from the vulnerabilities of 2G, 3G and 4G networks. Public key infrastructures (PKIs) have augmented that security by using certificates to allow entities within the network to identify each other and initiate secure communication, with X.509 certificates [1] the latest generation used in the PKI context.

The deployment of these security concepts in previous-generation mobile and IP-based networks makes them a helpful starting point for new approaches suited to 5G and today's fast-changing threat landscape.

1 X.509 is defined by the International Telecommunications Union's "Standardization Sector" (ITU-T), in ITU-T Study Group 17 and is based on ASN.1, another ITU-T standard.

In 5G networks, the role of authentication and encryption via digital certificates extends to all signaling traffic. Https uses mutual transport layer security (mTLS) along with digital certificates issued by the CSP's certificate authority (CA), and digital certificates are required to secure all application programming interface (API) calls across all layers of the CSP cloud network from virtualization to orchestration. In other words, certificates are used for authentication in a much larger proportion of total network communications in 5G than in previous "Gs" — across the RAN, core, transport and service layers.

That larger proportion of total network communications also includes slice-based use cases for everything from gaming to emergency services and the full range of connected devices, including the IoT. The security needs reach all the way from the end user into the nuts and bolts of each individual slice (see Figure 1).

Figure 1. The evolving role of digital certificates in telco networks

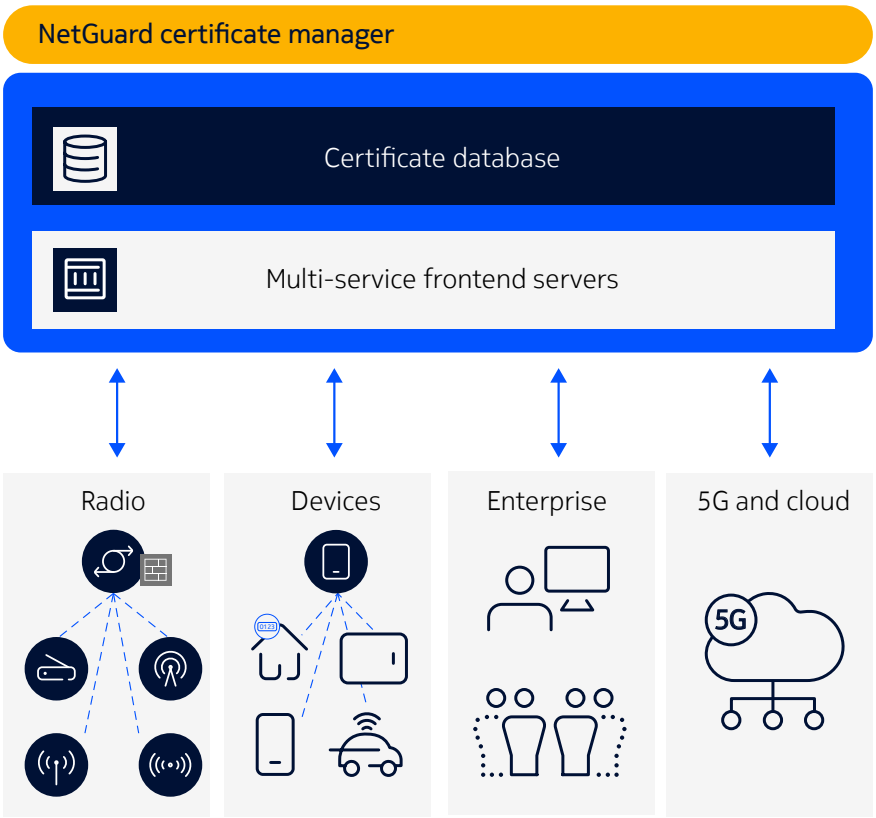
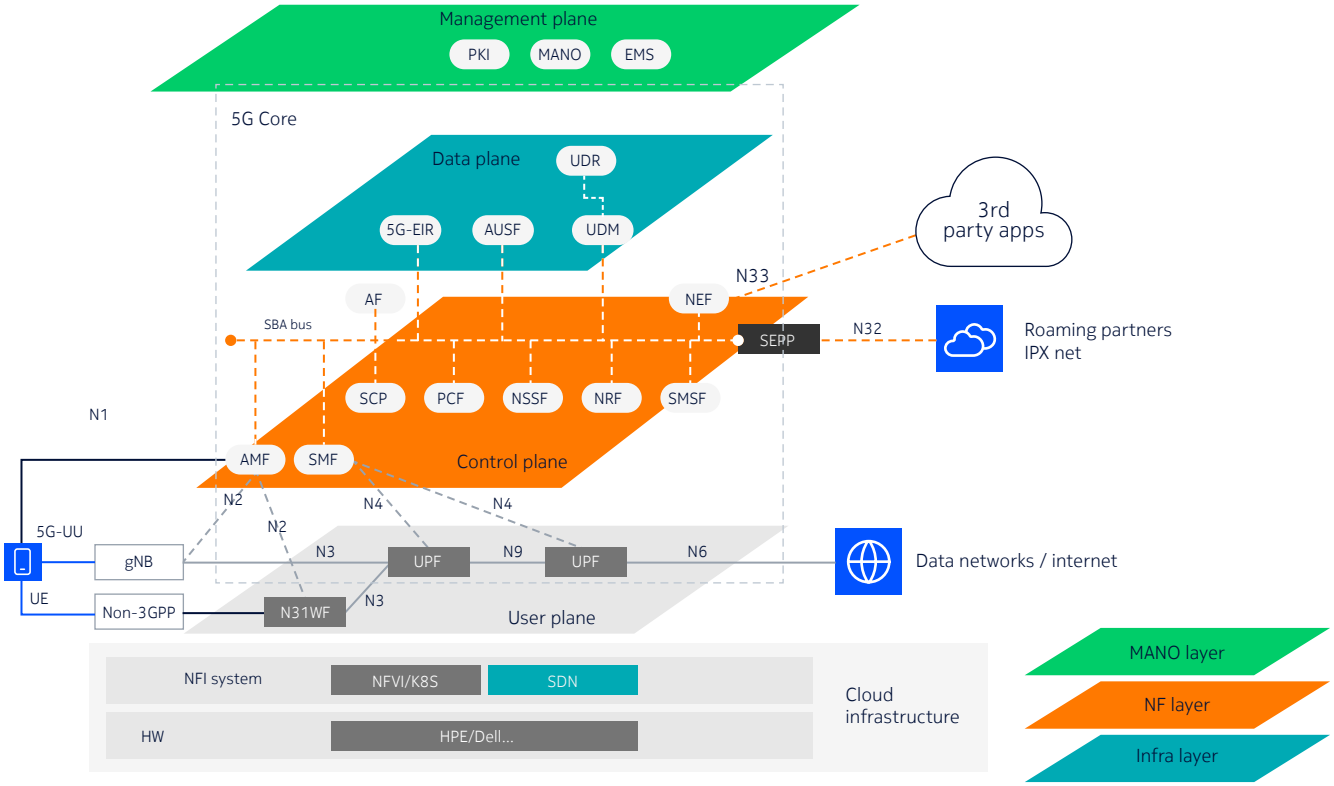


Figure 2 illustrates the same network-wide view of certificate requirements a little differently, showing how network function exposure domain and roaming interconnections, authentication and encryption are required to keep communications secure, from the management plane through the core and all the way to the service-based architecture (SBA).

Given the many functions within the network that require certificate-based security, certificates have a role in virtually every network domain. As an example, we can look at the data center infrastructure and fabric layer domain where 5G telco applications reside.

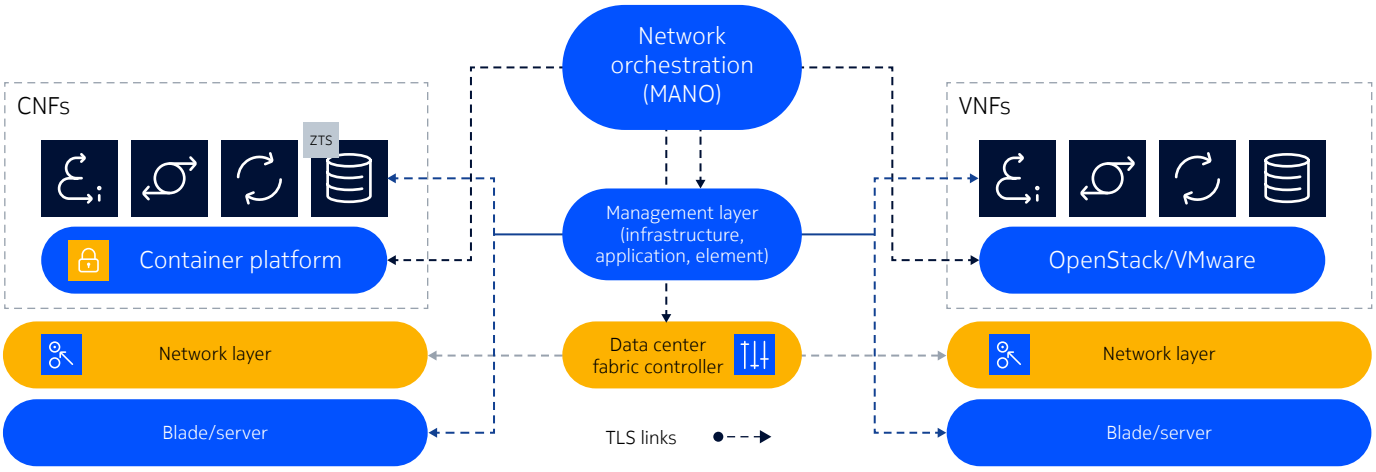
Figure 2. A domain-level view of where certificates are needed



This domain has three main components: the data center hardware, the hypervisor layer (CNF and/or VNF) and the data center fabric (the network layer) based on the GSMA Cloud Infrastructure Reference Model, as shown. (For descriptions of the various domains, see the Appendix to this paper.) Figure 3 shows in more detail the communication between the management and network orchestration (MANO) and operations domain, the administration and management (OAM) trust domain and the infrastructure domain.

The orchestration tool communicates with the management layer, the data center fabric and the hypervisor layer via application programming interface transport layer security (API TLS) communication as follows. This includes communication to manage and control the network layer (i.e., between the various components of the data center fabric) and intra-container communications where container network functions (CNFs) are involved. Both types of communication need to be secured via the use of digital certificates.

Figure 3. Communication between layers of the cloud data center



Intra- as well as inter-CNF communication is detailed in Kubernetes Istio.

The DSP approach to digital certificates

DSPs (and CSPs on the path to becoming them) can provision certificates by deploying a single central CA or several CAs, or by deploying several CA/PKI instances throughout the infrastructure. Either architecture will provide X.509 digital certificates for all components in the network across all domains.

The choice between single or multiple CA instances will depend on a DSP's infrastructure, security and compliance requirements, and on the need for logical and physical separation of the different services that will use certificates from the CAs. A 5G network slice providing a mission-critical service, for example, might require maximum separation, which would call for a separate CA instance for that slice only. If needed, configuring several root CAs separated from each other can give each service or domain its own root CA unrelated to any other root CA in the system, preventing cross-certification.

But in most cases, logical or physical separation of the CA solution is not required: a single CA can support multiple roots of trust for the network and its technologies instead of needing a CA for every root. Whatever the architecture, establishing the right configuration and managing the certificates prompts questions about what needs to be planned for in the immediate term, impacts on the 5G business strategy and more. The answers often call for expertise that many CSPs don't have.

Nokia 5G Security Consulting Services can provide that expertise where required. With more than 15 years of experience in security operations, Nokia security consultants have deep technical knowledge and real-world experience with end-to-end 5G security to support analytics, automation and customization. Nokia's consulting services are vendor-agnostic, aimed at optimizing architecture designs for each CSP environment, informed by significant experience in test and verification labs as well as with hundreds of clients.

Conclusion

As demonstrated in this paper, the need for digital certificates to authenticate entities and encrypt and obscure data is pervasive throughout 5G networks due to the complexity, dynamism and openness that come from fundamental virtualization, multi-layer interactions and the delivery of slice-based services.

Architecting for and managing that extent of digital certificate use may be daunting for many CSPs who are used to dealing with certificates only for securing base stations and IPSec. Nokia Security Consulting Services has extensive expertise to help CSPs meet their new certificate-based security requirements and seize the 5G opportunity at the same time.

For more information on how to successfully incorporate widespread digital certificates into a 5G network environment, and for further perspective on 5G security, contact Nokia Security Consulting Services.

Appendix: Certificate trust domains

Digital certificates have applications across the full range of CSP 5G network domains:

Radio access domain

This domain consists of base stations, Wi-Fi access points and transport connections to the core. It is based on the standardization by 3GPP and includes all type of base stations for 2G, 3G, 4G and 5G regardless of size and deployment mode. Macro, small cell and femto cells are deployed as bare metal (single system deployment) or in a distributed deployment. This domain is specified in 3GPP TS33.401, TS33.310 and TS33.210. Wi-Fi access points are defined as non-3GPP components.

Core domain

The 5G core is deployed as a cloud infrastructure with virtualized network function (VNF) workloads deployed on virtual machine functions and containerized or cloud-native network functions.

This means a fully automated and orchestrated VNF/CNF 5G core network requires several layers for the overall orchestration tool, the data center infrastructure and the data center fabric networking layer (taking the ESTI MANO architecture as a reference).

The core network for a mobile service is only one domain to be orchestrated. The data center orchestration and infrastructure layers could therefore be regarded as separate domains by themselves. However, for the sake of simplicity, all these layers are considered as a single core network in this document.

The following domains provide a good representation of the key areas for the deployment of digital certificates issued and managed by the telco service provider:

5G service-based architecture (SBA)

The SBA primarily covers the standardized interfaces between 5G network services, specifically the signaling layer. Also referred to as service-based interfaces (SBI) in 3GPP TS33.501.

Telco cloud orchestration layer

The cloud orchestration layer is the all the communications between the orchestration tool (using the ESTI MANO architecture as a reference) and the data center infrastructure, including communications between the network services applications.

The orchestration of this layer is not limited to core with its own tool, but may also be serviced by an overreaching tool or a cross-domain orchestrator.

Telco cloud data center fabric domain

The data center fabric network layer provides connectivity and communications within the core domain or to other domains. This domain includes both the virtual overlay network and the underlay physical networking layer within the data center.

Telco cloud data center infrastructure

This domain includes the communication between the different hypervisor layers. VNF and/or CNF is needed to provide the layer on which the telco service applications run.

National and international roaming interconnection

Based on 3GPP TS33.501, the security edge protection proxy (SEPP) operates at the edge of the source and destination networks to handle end-to-end core network interconnection security. The confidentiality and/or integrity of message elements is managed between two SEPPs of the source and destination public land mobile network (PLMN).[2] This traffic between the H-PLNM (home PLMN) and V-PLNM (visiting PLMN) is encrypted using https and digital certificates for authentication.

Exposure network function domain

This domain sits between the network exposure function (NEF) and external third-party applications. The purpose of the NEF is to support external exposure of capabilities of the 5G network functions to applications, which interact with the relevant network functions via the NEF (using the N33 interface) as defined by 3GPP TS33.501.

2 3GPP TS33.501 – section 5.9.3 https://www.3gpp.org/ftp/Specs/archive/33_series/33.501/33501-h20.zip

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID212562 (May)